[Patent Document]

Japanese Patent Laid Open
No. 2003-303055

Disk storage system having disk arrays connected with disk adaptors through switches

Hitachi, Ltd.

Inventor(s):Tanaka, Katsuya ;Fujimoto, Kazuhisa Application No. 10/212882, Filed 20020807, A1 Published 20031009 Abstract:

A disk storage system has high throughput between a disk adapter of a disk controller and a disk array. The disk adapter of the disk controller is connected to the disk array through switches. Data on a channel between the switch and a RAID group is multiplexed in the switch to be transferred onto a channel between the switch and the disk adapter and data on the channel between the switch and the disk adapter is demultiplexed in the switch to be transferred onto the channel between the switch and the RAID group. A data transfer rate on the channel between the disk adapter and the switch is made higher than that on the channel.

US.Class: 711114 711154

ディスクアダプタとディスクアレイをスイッチをかして接続したディスク装置

(19) 日本国金田(1 b)

€ 撒 ধ 罪 华 噩 (12) **(2)**

特開2003-303055 (P2003-303055A) (11)特許出國公開番号

平成16年10月24日(2003.10.24)	子子		12 A	0
平成16年	6	3 0	3.0	54
(43)公開日	3/08	3		
	F1	;		
	欧罗尼号 301			0 4 0
	98/			

G06F 3/ (51) Int C.

(4) II OL

(54) 「発明の名称」 ディスクアグプクとディスクアレイをスイッチを介して接続したディスク装置

(57) (現地)

【際图】 ディスクコントローラのディスクアダプタと ディスクアレイ間のスルーブットが高いディスク装置を 協供することにある。

· ダブタ (DKA) とディスクアレイ (DA) をスイッチ (SW1, SW スイッチ (S#1) とディスクアダプタ (DKA) 間のチャネル (D 01) 上のデータをスイッチ (SW1)において迎多皿化してス ータをスイッチ(S#I)において多皿化してスイッチ(S#I) 【解袂年段】 ディスクコントローラ(DKC)のディスク Dグループ(RI)間のチャネル(DI1, DI2, DI3, DI4)上のデ イッチ(SNI)とRAIDグループ(NI)間のチャネル(DII, 2. SW3, SW4) を介して接続する。スイッチ (SW1) とRA I とディスクアダプタ (DKA) 間のチャネル (D01) に転送し、 012, 013, 014)に低送する。ディスクアダプタ (DKA) とス イッチ (S#1)間のチャネル (D01) 上のデータ転送速度を、 チャネル(DI1,D12,D13,D14)のデータ転送選取より高く

O

ディスクアダプタとディスクアレイをスイッチを介して接続したディスク装置

ß

2

特開2003-3030

S 3 0 က - 30 特開2003

(特許請求の範囲)

からなり、前紀ディスクコントローラはチャネルアダプ ディスクコントローラとディスクアレイ タとキャッシュメモリとディスクアダプタを有するディ [請求項1]

前配ディスクアダプタと前配ディスクアレイを、パッフ アメモリを有するスイッチを介して接続し、 スク装置において、

前配スイッチは、前配ディスクアダプタが接続されたポ が接続された各ポートとの間でのポート間の接続の切り 換えを、入力されたフレーム毎に、敵フレーム内の送信 **ートと前記ディスクアレイを構成するディスクドライブ** 先情報にしたがって行うことを特徴とするディスク藝

アレイからなり、前配ディスクコントローラはチャネル ディスクコントローラと複数のディスク アダプタとキャッシュメモリとディスクアダプタを有す るディスク装置において、 [開水項2]

前記ディスクアレイはループ状に接続した複数のディス クドライブからなり、前紀ディスクアダプタと前紀複数 のディスクアレイとをパッファメモリを有するスイッチ を介して接続し、

前配ディスクアダプタと前配スイッチ間のチャネル当り データ転送速度を、前配スイッチと前配複数のディスク アレイ間のチャネル当りデータ転送速度より高く設定 前記スイッチは、前記ディスクアダプタが接続されたポ 一トと前配複数のディスクアレイが接続された各ポート との間でのポート間の接続の切り換えを、入力されたフ レーム毎に、散フレーム内の送信先情報にしたがって行 うことを特徴とするディスク装置。

|群求項3] ディスクコントローラとディスクアレイ: からなり、前起ディスクコントローラはチャネルアダプ **タとキャッシュメモリとディスクアダプタを有するディ** スク装置において、 前配ディスクアダプタと前配ディスクアレイを、パッフ アメモリを有するスイッチを介して接続し

可一のスイッチに接続したディスクドライブの組み合わ せでRAIDグループを構成し、

り換えを、入力されたフレーム毎に、放フレーム内の送 前配スイッチは、前配ディスクアダプタが接続されたポ 間先情報にしたがって行うことを特徴とするディスク藝 前紀ディスクアダプタと前紀スイッチ間のチャネル当り データ転送速度を、前記スイッチと前配ディスクアレイ ートと前紀RAIDグルーブを構成するディスクドライ ブが接続された各ポートとの間でのポート間の接続の切 間のチャネル当りデータ転送速度より高く設定し、

育1のディスクコントローラは第1のチャネルアダプタ 【酢求項4】 第1のディスクコントローラと第2のデ ィスクコントローラと複数のディスクアレイからなり、

と第1のキャッシュメモリと第1のディスクアダプタモ

と第2のキャッシュメモリと第2のディスクアダプタを 第2のディスクコントローラは第2のチャネルアダプタ 有するディスク装置において、

第1のディスクアダプタと前配複数のディスクアレイと をパッファメモリを有する第1のスイッチを介して接続 し、且つ第2のディスクアダプタと前配複数のディスク アレイとをパッファメモリを有する第2のスイッチを介 ダプタを接続し、第2のスイッチと第1のディスクアダ して接続し、さらに第1のスイッチと第2のディスクア ブタを接続し、

第1のディスクアダプタと第1のスイッチ間、および算 りデータ転送速度を第1のスイッチと前配複数のディス クアレイ間のチャネル当りデータ転送選段より高く設定 2のディスクアダプタと第1のスイッチ間のチャネル当 2

第2のディスクアダプタと第2のスイッチ間、および第 1のディスクアダプタと第2のスイッチ間のチャネル当 りデータ転送速度を第2のスイッチと前配複数のディス クアレイ間のチャネル当りデータ転送強度より高く設定 2

第1のスイッチは、第1のディスクアダプタまたは第2 15 イスクアレイが接続された各ポートとの間でのポート間 の接続の切り換えを、入力されたフレーム年に、鼓フレ のディスクアダプタが接続されたポートと前配複数のデ **- ム内の送信先情報にしたがって行い**

第2のスイッチは、第1のディスクアダプタまたは第2 のディスクアダプタが接続されたポートと前配複数のデ **イスクアレイが接続された各ポートとの凹でのポート**固 の接続の切り換えを、入力されたフレーム年に、鼓フレ **--ム内の送信先情報にしたがって行うことを特徴とする** ディスク装置。 8

「酵求項5】 第1のディスクコントローラと第2のデ 育1のディスクコントローラは第1のチャネルアダプタ と第1のキャッシュメモリと第1のディスクアダプタを 35 イスクコントローラと複数のディスクアレイからなり、

と第2のキャッシュメモリと第2のディスクアダプタを 第2のディスクコントローラは第2のチャネルアダプタ 有するディスク装置において、 \$

をパッファメモリを有する第1のスイッチを介して接続 アレイとをバッファメモリを有する第2のスイッチを介 第1のディスクアダプタと前配複数のディスクアレイと し、且つ第2のディスクアダプタと前配複数のディスク ダブタを接続し、第2のスイッチと第1のディスクアダ して接続し、さらに第1のスイッチと第2のディスクア ş

第1のディスクアダプタと第1のスイッチ間、および第 2のディスクアダプタと第1のスイッチ間のチャネル当 ブタを接続し、 20

ディスクアダブタとディスクアレイをスイッチ を介して接続したディスク装置

S

りデータ 転送速度を算1のスイッチと前配複数のディス クアレイ間のチャネル当りデータ転送遊取より高く設定

1 のディスクアダプタと第2のスイッチ間のチャネル当 クアレイ間のチャネル当りデータ転送速度より高く設定 第2のディスクアダプタと第2のスイッチ間、および第 りデータ転送速度を算2のスイッチと前配複数のディス

ダブタと第2のスイッチ固を被脱したチャネルと同等の ダブタと抑1のスイッチ間を接続したチャネルと同等の 類1のスイッチと類2のスイッチを、類1のディスクア データ転送速度を有するチャネルと、第2のディスグア 第1のスイッチは、第1のディスクアダプタまたは第2 のディスクアダプタまたは第2のスイッチが接続された ポートと前記複数のディスクアレイが接続された各ポー トとの間でのポート間の接続の切り換えを、入力された フレーム毎に、酸フレーム内の送信先情報にしたがって データ転送速度を有するチャネルと、を介して接続し、

第2のスイッチは、第1のディスクアダプタまたは第2 トとの間でのポート間の接続の切り換えを、入力された フレーム毎に、抜フレーム内の送信先情報にしたがって ポートと前配複数のディスクアレイが接続された各ポー のディスクアダプタまたは第1のスイッチが接続された **行うことを特徴とするディスク模団**

値配ディスクアレイからのデータ肌み出し時には、前記 前妃スイッチにおいて多瓜化して前記ディスクアダプタ ディスクアレイから付記スイッチに転送されるデータを 母記載のディスク被回において、

値配スイッチにおいて逆多重化して値配ディスクアレイ 位配ディスクアレイへのデータ母き込み時には、 紅配デ イスクアダプタから仰配スイッチに依送されるデータを に転送することを特徴とするディスク模型。

[請求項7] 請求項1乃至請求項5のいずれかの請求 ディスクアダブタからディスクアレイへのデータ告き込 み時に、前配ディスクアダプタは、前配ポート間の接続 の切り替えが周期的に行われるように、送出するフレー 項記載のディスク装配において、

ディスクアレイからディスクアダプタへのデータ観み出 し時に、包括スイッチは、ラウンドロピン方式により前 妃ポート間の彼彼を切り替えることを特徴とするディス ムに送伯先信頼を設定し、

る比、と同程度に設定することを仲徴とするディスク藝 イッチ間のチャネル当りデータ転送꿟取の、スイッチと ディスクアレイ図のチャネル当りデータ気送遊度に対す **えるポート数を、ディスクアダプタとス 酢水項7配紙のディスク数倒において、** 国歴色にわり [原水斑8]

顔求項1乃至醇求項5のいずれかの簡求 [四米四]

ープルで接続し、前配スイッチと前配ディスクアレイ関 をメタルケーブルで接続することを特徴とするディスク **前起ディスクアダプタと前配スイッチ間を光ファイパケ** 項記載のディスク装置において、

[発明の詳細な説明]

(0001)

テムにおける2次記憶装置に関し、特に入出カデータ転 [発明の属する技術分野] 本発明は、コンピュータシス 送性能が高いディスク装置に関する、

は、CPU(中央処理装置)が必要とするデータは2次 【従来の技術】現在のコンピュータシステムにおいて

配徴装置に保存され、CPUなどが必要とするときに応 じて2次配管装置に対してデータの音き込みおよび脱み 出しを行う。この2次記憶装置としては、一般に不揮発 **高度情報化に伴い、コンピュータシステムにおいて、こ** な妃篋媒体が使用され、代表的なものとして磁気ディス ク装団や、光ディスクなどのディスク装団がある。近年 の種の2次記憶装置の高性館化が要求されている。 20

[0003] 図9に、従来のディスク装置のブロック図 を示す。図9において、ディスク装置はディスクコント ローラDKCとディスクアレイDAで構成される。ディ スクコントローラDKCは、上位側CPU(図示せず) とディスク装置を接続するチャネルアダプタCHAと、

ディスクアレイDAに対して脱みむきするデータを一時 保存するキャッシュメモリCMと、ディスクコントロー

HAはC1、C2、C3、C4の4本のチャネルでCP ラDKCとディスクアレイDAを接続するディスクアダ シュメモリCMとディスクアダプタDKAは、パスまた はスイッチで相互接続されている。チャネルアダプタC Uと接続している。ディスクアダプタDKAはD1、D 2、D3、D4の4本のチャネルでディスクアレイと接 プタDKAからなる。チャネルアダプタCHAとキャッ 嵌している。ここでディスクアレイDAはディスクグル -JR1, R2, R3, R4からなる。ディスクアレイ DAにおいてRAIDシステムを構築する場合は、R

1. R2. R3. R4がそれぞれRAIDグループを構 [0004] チャネルC1、C2、C3、C4から入力 された母き込みデータは、キャッシュメモリCMに抜デ

ヤネルにはプロック単位に分割されたデータを、残りの 内に眩当データの有無を聞べる。有る場合は、キャッシ 一クを母き込むと同時に、散データをプロックサイズ単 位に分割し、チャネルD1、D2、D3、D4の内3チ ディスクアダプタDKAからディスクアレイDAへ送信 する。データ酰み出し時は、先ずキャッシュメモリCM 1 チャネルは前記分割データから計算したパリティを、 S

ャッシュメモリCM内に無い場合にディスクアダプタD KAは、D1、D2、D3、D4を介してディスクアレ し、チャネルアダプタCHAを介して競み出しデータを CPUへ送信する。 この種の従来技術を第1の従来技 インフレーム, 98] (1998年) 第144 鬨から第 ュメモリCMからチャネルアダプタCHAを介してキャ ッシュメモリ内酰み出しデータをC P Uへ送信する。キ 術と呼ぶ。第1の従来技術として関連するディスク装置 は、例えば、日経BP社刊の「日経コンピュータ別師メ イDAからプロック単位に分割されたデータを脱み出 153頁に記載されているディスク装置がある。

[0005] ディスクアダプタとディスクアレイを、ス 4の交換デバイス」に開示されている。以下、敵公報に 記載の従来の技術を第2の従来技術と呼ぶ。第2の従来 の技術によれば、ディスクアレイに関連したパス本数と **制御ブロックを介して接続したディスク装置が、特開平** 6-19627号の「回転形配像装置」に開示されてい イッチを介して接続したディスク装団が、特関平 5 - 1 13122号の「マルチチャンネルデータおよびパリテ ディスクアダプタに関連したパス本数とを独立に散定で る。以下、鼓公報に記載の従来の技術を第3の従来技術 と呼ぶ。第3の従来技術によれば、ディスクアダプタと きる。ディスクアダプタとディスクアレイを、バッファ ディスクアレイ間のデータ転送速度を任意に設定でき、 ディスクの回転待ちの影響を低減できる。

ネルでは、現状でチャネル当りのデータ転送速度が1G エンドに比く高速化が難しいこと、である。 第2の理由 は、たとえディスクドライブが高速化したとしても、全 歩に伴い、1チャネル当りのデータ転送速度は年々増加 している。例えばディスク装置に使われるファイバチャ bpsから2Gbpsであるが、近い将来4Gbpsか れる。第1の理由は、ディスクドライブは機械部品を含 てのディスクドライブ毎に高速インターフェイスを搭載 するのは、多数のディスクドライブを有するディスク装 [発明が解決しようとする課題] ネットワーク技術の進 CPUとチャネルアダプタ間(以下フロントエンドと呼 **ぶ)のスールプットはこの高速化に従うことが予想され** (以下パックエンドと呼ぶ) のスルーブットは以下の理 由により、フロントエンドほど高速化されないと予想さ むので、電子、光楽子のみ高速化を行えば良いフロント る。ところが、ディスクアダプタとディスクアレイ間 510Gbpsへ萬滋化されることが予定されている。 賢の高コスト化を招くことである。

[0007] 第1の従来技術では、チャネルアダプタの トを多数、ディスクアダプタに散けることも考えられる チャネル当りのデータ転送速度を向上させても、フロン トエンドとパックエンドのスループット和幅により、デ ィスク装置の性値が向上しないという問題があった。ま た、パックエンドのスループット向上のために低速ポー

ディスクアダプタのボート数増加は傾御を複雑とす る。第2の従来技術では、ディスクアダプタとディスク アレイとの間にスイッチを適用することによりディスク 増設ポート数を増加させることができるが、チャネル当 りのデータ転送速度はディスクアレイのデータ転送速度 に倒限されるので、ディスクアダプタとディスクアレイ 聞のスループットが性能ネックになるという問題があっ た。 第3の従来技術は、ディスクの回転特ち時回の影響 を低減できる技術であり、フロントエンドとバックエン ドのスループット範疇は低減できないという問題があっ

タとディスクアレイ間ネットワークを有するディスク装 [0008] 本発明の目的は、ディスクアダプタとディ スクアレイ間のスループットが高いディスク装配を提供 することにある。本発明の他の目的は、ディスクアダブ 15 タとディスクアレイ配のスループットが苗へ、且つディ スクドライブ接続台数が多いディスク装置を提供するこ とにある。本発明のさらに他の目的は、信頼性が高いデ ィスクアレイを有するディスク藝囮を提供することであ る。本発明のさらに他の目的は、信頼性が高いディスク アダプタとディスクアレイ関ネットワークを有するディ スク装置を提供することにある。本発明のさらに他の目 的は、偕頼性およびスループットが高いディスクアダプ ディスクからの睨み出しおよびディスクへの母きこみを 高スループット化できるディスク装配を提供することに ある。本発明のさらに他の目的は、商スループットを植 **やできるディスク装置を提供することである。本発明の 母を提供することにある。本発明のさらに他の目的は、**

さらに他の目的は、高スループットで低コストなディス ク装置を提供することである。 22

からなり、ディスクコントローラはチャネルアダプタと キャッシュメモリとディスクアダプタを有するディスク 抜囮であり、ディスクアダプタとディスクアレイを、パ ッファメモリを有するスイッチを介して接続し、ディス クアダプタとスイッチ間のチャネル当りデータ転送速度 め、本発明は、ディスクコントローラとディスクアレイ 【喋盟を解決するための手段】上配目的を遊成するた

を、スイッチとディスクアレイ間のチャネル当りデータ 気送強度より高く設定し、スイッチは、ディスクアダブ スクドライブが接続された各ポートとの間でのポート間

タが後続されたポートとディスクアレイを格成するディ

の接続の切り換えを、入力されたフレーム年に、放フレ **一ム内の送信先情報にしたがって行っている。また、前** 紀ディスクアレイはループ状に接続した複数のディスク ドライブからなり、恒配ディスクアダプタと値配複数の ディスクアレイとをパッファメモリを有するスイッチを 介して接続し、ディスクアダプタとスイッチ間のチャネ ル当りデータ転送遊度を、スイッチと複数のディスクア レイ節のチャネル当りゲータ院秘涵技より箱へ設定し、 \$ 20

47.

た、他紀ディスクアダプタと位紀ディスクアレイを、パ スイッチは、ディスクアダプタが接破されたポートと複 数のディスクアレイが接続された各ポートとの間でのポ ッファメモリを有するスイッチを介して接載し、同一の スイッチに接続したディスクドライブの組み合わせでR 間のチャネル当りデータ転送速度を、スイッチとディス クアレイ間のチャネル当りデータ転送強度より高く設定 数フレーム内の送信先信報にしたがって行っている。 虫 AIDグループを構成し、ディスクアダプタとスイッチ し、スイッチは、ディスクアダプタが接続されたポート とRAIDグループを傾成するディスクドライブが接続 **一ト間の後戌の切り換えを、入力されたフレーム毎に、**

報にしたがって行っている。また、第1のディスクコン のチャネルアダプタと第1のキャッシュメモリと第1の は類2のチャネルアダプタと類2のキャッシュメモリと トローラと節2のディスクコントローラと複数のディス を、入力されたフレーム低に、鼓フレーム内の送信先情 クアレイからなり、抑1のディスクコントローラは第1 ディスクアダプタを有し、第2のディスクコントローラ された各ポートとの間でのポート間の接続の切り換え

用1のディスクアダプタと前配複数のディスクアレイと をパッファメモリを有する第1のスイッチを介して接収 し、且つ第2のディスクアダプタと前配複数のディスク 訊2のディスクアダプタを有するディスク藝匠であり、

間のチャネル当りデータ転送速度を第2のスイッチと前 アレイとをパッファメモリを有する第2のスイッチを介 して接続し、さらに第1のスイッチと第2のディスクア ダブタを接続し、第2のスイッチと第1のディスクアダ プタを接続し、第2のディスクアダプタと第2のスイッ チ間、および第1のディスクアダプタと第2のスイッチ

配複数のディスクアレイ間のチャネル当りデータ転送速 敗より商く設定し、第1のスイッチは、第1のディスク アダプタまたは第2のディスクアダプタが接続されたポ との間でのポート間の接続の切り換えを、入力されたフ レーム毎に、数フレーム内の送信先情報にしたがって行 い、第2のスイッチは、第1のディスクアダプタまたは 第2のディスクアダプタが接破されたポートと前配複数 ト間の接続の切り換えを、入力されたフレーム毎に、厳 ートと印配複数のディスクアレイが接続された各ポート のディスクアレイが後載された各ポートとの間でのポー フレーム内の送信先情報にしたがって行っている。虫

と、第2のディスクアダプタと第1のスイッチ間を接続 したチャネルと同等のデータ転送速度を有するチャネル からのデータ競み出し時には、値配ディスクアレイから 前配スイッチに転送されるデータを前配スイッチにおい 上記算1のディスクアダプタと第2のスイッチ間を接続 したチャネルと同等のデータ転送遊底を有するチャネル て多皿化して前配ディスクアダプタに転送し、前配ディ と、を介して接続している。また、前配ディスクアレイ た、さらに、上配算1のスイッチと第2のスイッチを、

ッチにおいて逆多選化して前配ディスクアレイに転送す るようにしている。また、前紀ディスクアダプタから前 紀ディスクアレイへのデータ番き込み時に、前紀ディス クアダプタは、前記ポート間の接続の切り替えが行われ また、さらに、切り替えるポート数を、ディスクアダプ るように、送出するフレームに送信先情報を設定し、前 **記ディスクアレイから前記ディスクアダプタへのデータ 酰み出し時に、前記スイッチは、ラウンドロピン方式に** ダブタから前配スイッチに転送されるデータを前配スイ スクアレイへのデータ歯食込み時には、前配ディスクア より前記ポート間の接続を切り替えるようにしている。

镣幌し、前記スイッチと前記ディスクアレイ間をメタル [発明の実施の形態] 以下、図面を参照して本発明の実 ケーブルで検幌するようにしている。 [0010]

スクアダプタと前記スイッチ間を光ファイパケーブルで

に対する比、と同程度に散定している。また、前配ディ

タとスイッチ間のチャネル当りデータ転送速度の、スイ ッチとディスクアレイ間のチャネル当りデータ転送速度

施の形態であるディスク装置の構成を示す。 本実施の形 Cは、チャネルアダプタCHAと、キャッシュメモリC 娘のディスク装置は、ディスクコントローラDKCとデ ィスクアレイDAからなる。 ディスクコントローラDK 施の形態を詳細に説明する。図1に本発明の、第1の実

Mと、ディスクアダプタDKAからなる。チャネルアダ ð. C1, C2, C3およびC4は、チャネルアダプタ プタCHAは、上位CPU(図示せず)とディスクコン CHAがCPUと通信するチャネルである。 キャッシュ メモリCMは、本実施の形態のディスク装置が入出力す *እዕፓሃプタ*DKAは、チャネルD01、D02、D0 るデータを一時保存するメモリである。ディスクアダプ タDKAは、ディスクコントローラDKCとディスクア レイDAとがデータを送受値する際の慰御を行う。ディ トローラDKCとがデータを送受信する際の制御を行

ネルD01、D02、D03、D04上で全二重通信が ディスクアダプタDKAとディスクアレイDAは、チャ 3、D04を介して、ディスクアレイDAと接続する。 {0011} ここで、本実施の形態のディスク装置は、 可能である。

ッチSW1、SW2、SW3、SW4を介して接続して いる点に特徴がある。ディスクアレイDAは、ディスク DKAと接続する。同様に、ディスクグループR2はス イッチSW2を介して、ディスクグループR3はスイッ チSW3を介して、ディスクグループR4はスイッチS ディスクアダプタDKAとディスクアレイDAを、スイ グループR1, R2, R3, R4からなる。ディスクグ ループR 1は、スイッチSW1介してディスクアダプタ W4分して、それぞれディスクアダプタDKAと接続す

[0012] 本実施の形態のディスク装置においてRA 2

I Dシステムを構築する場合は、ディスクグループR

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

1, R2, R3, R4&, EnenRAID///-/2 スクグループへのデータ航み出しまたは杳き込み時のデ する。本実施の形態では、4個のディスクドライブでR AIDグループを構成しているが、RAIDグループを 構成するドライブ数を4個に限るものではない。各ディ である。チャネルC1、C2、C3、C4からディスク グループR 1へ番き込むためにCPUから送信されたデ ータは、ディスクアダプタDKAにおいてブロック単位 に分割されると同時に、駭ブロック単位に分割されたデ **一夕からパリティが生成される。 放ブロック単位に分割** されたデータと、生成されたパリティは、チャネルD0 1を通りスイッチSW1へ入力される。スイッチSW1 は、RAID制御に伴い、敢プロック単位に分割された データと、生成されたパリティとをルーティングし、チ 1. D12. D13. D14を介してディスクグループ る.ここでRIはRAIDレベル5のRAIDグループ ータの斑れを、ディスクグループR 1を例にして述べ ータ競み出し時は、ディスクアダプタDKAは、D1 ヤネルD11, D12, D13, D14へ分配する。

[0013] 図9に示した従来のディスク装置では、デ スイッチSW1通過後に別々のチャネルに分配される点 D3、D4上で、既にディスクアレイへの留き込みデー イスクアダプタDKAに接続したチャネルD1、D2、 タおよびパリティが別々のチャネルに分配されていた。 それに対し、本実施の形態のディスク装置においては、

スイッチSW1でシリアル化して、チャネルD01を介

して観み出しデータを受信する。

R1からブロック単位に分割されたデータを飲み出し、

であるスイッチの動作を、スイッチSW1を例にとり説 1、P2、P3、P4、P5は、全二直通信可能な入出 [0014] 次に、本実施の形態のディスク装置の特徴 明する。SW2~SW4の動作もSW1の動作と同様で カポートであり、ポート毎にパッファメモリを有してい る。スイッチSW1の内部構成を図2と図3に示す。値 D13、D14上を流れるデータは、フレーム単位で送 単のため、データの進行方向によりスイッチ動作を分け ある。図1に示すように、スイッチSW1は入出力ポー 受信され、かつデータは8B10B変数で符号化されて て説明する。また、チャネルD01、D11、D12、 トP1、P2、P3、P4、P5を有する。ポートP

5のクロスパスイッチであり、入力ポートin1、in - ムを入力し、ポートP2、P3、P4、P5から出力 する場合を示す。これはディスクアレイへの者を込み時 [0015] 図2は、ポートP1からブロック内のフレ のスイッチ動作に相当する。スイッチSW1は図2に示 すように、クロスパスイッチXSWと、スイッチコント ローラCTLからなる。クロスパスイッチXSWは5×

アメモリBM2と、バラレルシリアル変換装図PS2を out 2. out 3. out 4. out 5を有する。 ポ **一トP1から入力したフレームは、シリアルパラレル変 換装置SP1と、パッファメモリBM1と、8B10B 変換デコーダDEC1を経由し、スイッチコントローラ** て選ばれた場合は、入力したフレームは出力ポートou **CTLと入力ポート।n1へ入力される。スイッチコン** 掛かれた送信先アドレスを解脱し、クロスパスイッチ X 2, 1n3, in4, in5と、出力ポートout1, トローラCTLにおいて、入力フレームのヘッダ部分に SWを切り換える。例として、ポートP2が出力先とし t 2 と、8 B 1 O B変換エンコーダENC 2 と、パッフ

アメモリBM1、BM2はFIFO (First-In 紙由して、ポートP2から出力される。ここで、パッフ [0016] シリアルパラレル党徴被回SP1は、8B 10B符号化されたシリアルデータを10bit幅のパ First-out) メモリである。

ラレルデータに敷換し、ポートP1におけるデータ転送 に番き込む。8月10日デコーダDEC1は、クロスパ スイッチXSWの動作速度に同期して、10bltパラ 5. 8B10BエンコーダENC2は、クロスバスイッ び8B10B符号化し、10bitパラレルデータに変 数後、クロスパスイッチXSWの動作速度に同切してパ **速度の1/10の速度に同期してパッファメモリBM1** レルデータをパッファメモリBM1から散み出し、8B チXSWでスイッチされた8bitパラレルデータを再 ッファメモリBM2に毎き込む。パラレルシリアル変換 **複图PS2は、ポートP2におけるデータ転送速度の1** /10の遊度に同煩して、10511パラレルデータを は、ポートP1におけるデータ転送速度からポートP2 パッファメモリBM2から散み出し、シリアル化して、 10B毎号化して、8bitパラレルデータに変換す ポートP2から出力する。以上によりスイッチSW1 25

と、ポートP2、P3、P4、P5から出力されるフレ 一ムを示した図である。故形の凸はフレームが存在する 出力ポートP2、P3、P4、P5におけるデータ転送 ンシャルアクセスが行われており、フレーム長が一定で ある。図4では、入力ポートP1でのデータ転送遊度が 速度のm倍あるとする。従って、ポートP1におけるフ る。フレームは伝送するデータ容量に従ってそのフレー ム長が変化するが、ここではディスクアレイへのシーケ レームFb2の時回T1は、ボートP2からの出力時に [0017] 図4は、ポートP1へ入力するフレーム 時間、凹はフレームが存在していない時間を示してい におけるデータ転送速度へ速度変換する。 **\$**

[0018] 入力のデータ転送協収が強く、且つ出力の データ転送速度が遅い場合は、スイッチを周期的に切り ブットが低下する。フレームがスループットの低下無く 換えないと出力ポートのパッファメモリが溢れ、スルー T3へ伸びている。 にこでT3=EXT1である。

20

ば、フレームの簡敬無く、スループットの低下は危こら スイッチを迢迢するには、図4のように周期的に出力ポ 一トを切り換える必要がある。スイッチ切替えポート数 ない. T2≥T3はn≧mと同じある。つまり、ディス クアレイへのデータ音き込み時に、スイッチにおいてス イッチ間のチャネル当りデータ転送速度の、スイッチと で周期的に切り替えることにより逆多皿化し、ポートP をnとすると、スイッチ切替え周辺T2≒n×T1であ る (フレームの無い時間は無ねした) . T2≥T3なら ルーブット低下を起こさないための条件は、周期的に切 2、P3、P4、P5へ分配して出力する。スイッチを **周期的に切り飲える方法の一つは、スイッチに接続した** り替えるスイッチポート数nを、ディスクアダプタとス ディスクアレイ間のチャネル当りデータ転送函度に対す る比n、以上に設定することである。この条件が保たれ れば、スイッチSWIは、ポートPIから入力したデー タをパッファメモリにおいて遊政政役は、フレーム単位 る. RAIDのストライピング的抑に従えば、スイッチ ディスクグループをRAIDグループとすることであ は周期的に切り替わる。

らフレームを入力し、ポートP1から出力する場合を示 す。これはディスクアレイからの脱み出し時のスイッチ 解脱し、クロスパスイッチXSWを切り換える。図3の [0019] 図3は、ポートP2、P3、P4、P5か 動作に相当する。例えば、ポートP2から入力したフレ 由し、スイッチコントローラCTLと入力ポート!n2 場合は、ラウンドロピン方式によりクロスパススイッチ 2、 P3、 P4、 P5) に同時にフレームが届く。 これ **一ムは、シリアルバラレル変換装置SP2と、パッファ** メモリBM2と、8B10B変換デコーダDEC2を経 入力フレームのヘッダ部分に僻かれた送信先アドレスを はない。スイッチは、総当り的に入出力ポート間接税を フレームずつ出力ボート (P1) へ転汲する。このよう に、スイッチを勧当り的に切り替える方式を、ラウンド ら複数の入力フレームは同期して入力ポートに届く必要 フレームは出力ポートのロt1と、8B10B変換エン コーダENC1と、バッファメモリBM1と、バラレル 切り替えることにより、これら複数の入力フレームを1 により、結果的にスイッチは周期的に切り替わることに なる。なお、読み出し時においても、スイッチはフレー XSWを切り替えて、順番にポートP2、P3、P4、 ロビン(Round Robin)方式と呼ぶ。ラウンドロビン方式 へ入力される。スイッチコントローラCTLにおいて、 ム内送信先信頼に従って切り替わることに適いはない。 P 5から入力されるデータは全てポートP 1 〜出力す る. すなわち、脱み出し時は、複数の入力ボート (P

10B符号化されたシリアルデータを10blt幅のパ 50 [00.20]シリアルバラレル変換袋図SP2は、8B

ツリアル資数板原PS1を箱由して、ボートP1から出

ラレルデータに変換し、ポートP2におけるデータ転送 に雷き込む。8 B 1 0 B デコーダDEC2は、クロスパ 速度の1/10の速度に阿期してパッファメモリBM2 スイッチX S Wの動作速度に同期して、10bit/パラ レルデータをパッファメモリBM2から就み出し、8B 10日復号化して、8biじパラレルデータに変換す

5. 8B10BエンコーダENC1は、クロスパスイッ チX SWでスイッチされた 8 b i t パラレルデータを再 び8B10B符号化し、10b1tパラレルデータに変 校後、クロスパスイッチX S Wの動作遊度に同期してパ /10の速度に同期して、10bitパラレルデータを は、ポートP2におけるデータ転送適度からポートP1 ッファメモリBM1に告き込む。パラレルシリアル変換 装置PS1は、ポートP1におけるデータ転送速度の1 パッファメモリBM1から脱み出し、シリアル化して、 ポートP1から出力する。以上によりスイッチSW1 におけるデータ転送速度へ速度変換する。

[0021] 図5は、ポートP2、P3、P4、P5へ ムを示した図である。彼形の凸はフレームが存在する時 が政化するが、ここではディスクアレイへのシーケンシ る。図5では、入力ポートP1でのデータ転送速度が出 カポートP2, P3, P4, P5におけるデータ転送選 度のm倍あるとする。従って、ポートP5におけるフレ ムFe2, Fe3, Fe4, Fe5をポートP1から出 フレームは伝送するデータ容量に従ってそのフレーム長 **一ムFe5の時間T4は、ポートP1からの出力時にT** 力するのにかかる時間をT6とする。スイッチ切り替え ポート数をnとすると、T6ちn×T5である (フレー ムの無い時間は無視した)。スイッチにおいて輻輳によ るスループット低下を防止するためには、T6≤T4と 入力するフレームと、ポートP1から出力されるフレー 5~種んでいる。ここでT4=m×T5である。フレー 間、凹はフレームが存在していない時間を示している。 ャルアクセスが行われており、フレーム長が一定であ する必要がある。T6≤T4は,n≤mと同じある。

[0022] つまり、ディスクアレイからのデータ観み 出し時に、スイッチにおいてスループット低下を起こさ ないための条件は、周期的に切り替えるスイッチポート 数nを、ディスクアダプタとスイッチ間のチャネル当り データ転送速度の、スイッチとディスクアレイ間のチャ る、ことである。この条件が保たれれば、スイッチSw タをパッファメモリにおいて速度変換し、フレーム単位 で周期的に切り替えることにより多異化し、ポートP1 ト数を、ディスクアダプタとスイッチ間のチャネル当り データ転送速度の、スイッチとディスクアレイ間のチャ へ出力する。よって、ディスクアレイへの警ぎ込みおよ びディスクアレイからの競み出しを高スループット化す 1は、ポートP2、P3、P4、P5から入力したデー るためには、nim、つまり、周期的に切り えるポー ネル当りデータ転送速度に対する比m、以下に設定す

ネル当りデータ転送速度に対する比、と同程度に設定す

ì

i

S

特別2003-30305

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

ればよいことが分かる。

ので、実効的なスルーブットは2Gbps×4=8Gb [0023] 例えば、ディスクアダプタとスイッチ間の 4Gbpsのチャネル1本で接続し、スイッチとディス クアレイ間を16bpsのチャネル4本で接続する。ま た、ディスクアダプタとスイッチ間の10Gbpsのチ ャネル1本で接続し、スイッチとディスクアレイ間を2 Gbpsのチャネル4本で接続する。この場合、スイッ チ入出カポート間でスループットのバランスが取れない

[0024] 以上より、スイッチSW1において選度変 でも、チャネルD01、D02、D03、D04でのデ できる。本実施の形態のディスク装置におけるデータ転 送方式としては、ファイバチャネルやインフィニパンド D12、D13、D14上のデータ転送速度が低速 タDKAとディスクアレイDA間のスループットを向上 ータ転送速度は高速にできる。つまり、ディスクアダプ 数と多頂化、逆多国化が行われるので、チャネルD1

[0025] 図6は、第1の実施の形態のディスク装置 において、ディスクドライブの増設方法を示した図であ 能性がある。そこでスイッチSW1では、ディスクアダ プタDKAとの間に、新規チャネルD05を増設してい 図6では図1に対して、ディスクグループR5とR め、スイッチSW1とSW2としてボート数の多いスイ ディスクアダプタ倒とのスループットバランスが協れる る。また、スイッチSW2の場合は新規チャネルを増設 せず、チャネルD02の信号伝送速度を増加させること ル8本で接続し、ディスクアダプタとスイッチ間を4G ル8本で接続し、ディスクアダプタとスイッチ間を10 Gbpsのチャネル1本で接続する。このように、本実 ので、スイッチの速度変換機能が有効に働かなくなる可 は、スイッチとディスクアレイ間を1Gbpsのチャネ は、スイッチとディスクアレイ間を1Gbpsのチャネ 箱の形態のディスク装置は、スイッチのポート数に広じ で、ディスクアダプタ個とディスクアレイ似のスループ て、ディスクドライブを増設可能である。 このディスク ドライブ増設方法は、1ポート当たりに接続できるドラ [0026] 図7に本発明の、第2の実施の形態である イブ数が少ないATA (AT Attachment) ッチを使用している。ディスクドライブを増散すると、 スイッチのディスクアレイ匈のスループットが増加し、 ットパランスを取っている。例えばスイッチSW1で 6が増設されている。ディスクドライブを増設するた bpsのチャネル2本で核焼する。スイッチSW2で 方式ディスクドライブを増設するのに適用できる。

しと呼ぶ)がめる。

ィスク装配は、ディスクコントローラDKCと、4個の F1X97V1DA1, DA2, DA3, DA4#54 る。ディスクコントローラDKCは、チャネルアダプタ CHA、キャッシュメモリCM、ディスクアダプタDK Aからなる。ディスクアレイDA1とディスクアダプタ DKAは、チャネルD01とスイッチSW1を介して接 睨する。同様に、ディスクアレイDA2はチャネルD0 2とスイッチSW2を介して、ディスクアレイDA3は チャネルD03とスイッチSW3を介して、ディスクア て、それぞれディスクアダプタDKAと接破する。スイ ッチSW1、SW2、SW3とSW4は、第1の収筋の 形態と同様に速度変換と多皿化、逆多田化を行うスイッ チとして機能する。本実施の形態におけるディスクアダ JADKAŁ, ZイッチSW1, SW2, SW3, SW 4*L.* 74*A*97*V*4*D*A1, *D*A2, *D*A3, *D*A4 との間のデータ転送方式は、ファイパチャネルを使用し ている。スイッチSW1、SW2、SW3、SW4はフ ァイバチャネルスイッチである。

レイDA4はチャネルD04とスイッチSW4を介し

[0027] 本奥施の形像におけるディスクアレイの構 板を、ディスクアレイDA1を倒に述べる。ディスクア と、D12上に接続した4個のディスクからなるディス クアレイと、D13上に依殺した4個のディスクからな るディスクアレイと、D14上に依拠した4個のディス を例にとると、ディスクドライブDK1、DK2、DK イバチャネルアーピトレイテッドループ (以下FC-A りからなるディスクアレイ、からなる。チャネルロ11 のように、多数のドライブを一つのチャネル上に接扱し レイDA1、DA2、DA3、DA4は、 **両袋のドライ** てディスクドライブにアクセスする方法としては、ファ ブ術成である。ディスクアレイDA1は、チャネルD1 1 上に接続した 4 個のディスクからなるディスクアレイ 3. DK4が、チャネルD11上に接続されている。 52

クドライプDK1、DK2、DK3、DK4の協模形盤 を例として示す。各ディスクドライブの入出力ポートね よびスイッチSW1の入出力ポートは、送信機T×と受 10に示すように、各ドライブの入出力ポートおよびス ある。各ドライブの入出力ポートはファイバチャネルの [0028] 図10に、FC-ALの依拠形態をディス 個機R×を有する。FC-ALの接続形態は、例えば図 イッチの入出力ポートをループ状に接続するトポロジで NI (Node Loop) ポートとして協信する。N しポートとは、ループ動作をする装団(ここではディス クドライブ) のポートである。スイッチSW1のディス クアレイDA1接続個入出力ポートは、ファイバチャネ する。FLポートとは、FC-ALを被破可能なスイッ ルのFL(Fabric Loop) ポートとして機能 チのポートである。FLポートを有するループは、ファ イバチャネルのパブリックループとして機能するので、 35

20

スクアレイ部分の構成方法が異なる。本実施の形態のデ

ディスク装団の構成を示す。本実施の形態のディスク装

置は、第1の実施の形態のディスク装置に対して、ディ

ディスクアダプタとディスクアレイをスイッチ をかして接続したディスク装置

1. DK 2. DK 3. DK 4 tt. 21 " FSW 1 2 LT ブとなる。パブリックループとは、ループ上のディスク ドライブが、スイッチを介してルーブ外のボートと通信 チャネルD01を介してディスクアダプタDKAと通信 可値である。以上、チャネルD11の接続形態を例に説 明したが、チャネルD12、D13、D14でも同様で 本実施の形倣では、4個のディスクドライブでRAID チャネルD11が形成するFC-ALはパブリックルー ある.本実施の形態のディスク裝置においてRAIDシ グループを構成しているが、RAIDグループを構成す 2、R3、R4を、それぞれRAIDグループとする。 可能なループである。よって、ディスクドライプDK ステムを構築する場合は、ディスクグループR1、R

14上には、それぞれ最大126台までのディスクドラ 1, D12, D13, D14において, それぞれFC-ALを用いてディスクドライブを接続している。FC— ALの仕様から、チャネルD11、D12、D13、D を、チャネルD11, D12, D13, D14の媒体と [0029] 本奥施の形態においては、チャネルロ1 イブが接続可値である。また、チャネルD01、D0 2、D03、D04の媒体として光ファイパケーブル るドライブ数を4個に取るものではない。 してメタルケーブルを用いる。

スク数団においては、ディスグドライブをFC-ALで **散焼しているので、スイッチのボート当りに接続できる** ドライブ台数が増加できる。つまり、ディスク装置当り の記憶容量を増加させる効果がある。また、ディスクド ライブをメタルケーブルで接破することにより、ディス クドライブ毎に高価な光インターフェイスを装備する必 **要がなくなるので、ディスクドライブのコストを下げる** [0030]以上説明したように、本実施の形態のディ

団は、ディスクコントローラとスイッチを二重化した点 [0031] 図8に本発明の、第3の実施の形態である ディスク装団の構成を示す。本実施の形態のディスク装 に怜徴がある。 本実施の形態において、ディスクアダブ

スイッチSW1、SW2と、ディスクアレイDA1から て做値する。ディスクコントローラDKC1は、チャネ なる。スイッチSW1とSW2は、第1の実施の形態と 同様に斑斑斑粒と多皿化、逆多皿化を行うスイッチとし ディスクアレイDA1との間のデータ転送方式は、ファ イパチャネルを使用している。 本実施の形態のディスク タDKA1, DKA2と、スイッチSW1, SW2と、 被配は、ディスクコントローラDKC1、DKC2と、

ッチ SW 2 をチャネルD 2 a で接続し、ディスクアダプ し、ディスクアダプタDKA2とスイッチSW1をチャ タDKA1とスイッチSW2をチャネルD1bで接続 ネルD2bで接続する。

- [0032] ディスクアレイDA1を構成するディスク ネルD 1 1 およびD 2 1 の両チャネルと接続する。ディ スクアレイDA1は、チャネルD11とD21に接続し 30 F 5 1 7 DK 1, DK 2, DK 3, DK 4 は, チャ ドライブは、入出力ポートを2個有する。例えば、ディ
- た4個のディスクからなるディスクアレイと、D12と るディスクアレイと、D14とD24に接続した4個の ディスクからなるディスクアレイ、からなる。チャネル D11, D12, D13, D14, D21, D22, D D22に接続した4個のディスクからなるディスクアレ イと、D13とD23に接続した4個のディスクからな 23、D24は、FC-ALでディスクドライブを接続

3、DK4の接続形態を例として示す。各ディスクドラ [0033]図11に本実施の形態におけるFC-AL の接続形態を、ディスクドライブDK1、DK2、DK イブは、それぞれNLポートを2個有する。各ディスク ドライブの入出力ポートおよびスイッチSW1、SW2 の入出力ポートは、送信機T×と受信機R×を有する。

- 入出力ポートは、FLポートである。チャネルロ11に より、スイッチSW1、ディスクドライプDK1、DK 2、DK3、DK4をループ状に接続する。同様にチャ ネルD21により、スイッチSW2、ディスクドライブ スイッチSW1、SW2のディスクアレイDA1接模側 DK1, DK2, DK3, DK4をループ状に接続す
 - る。これら2個のループは、ファイバチャネルのパブリ DK3、DK4は、スイッチSW1またはSW2を介し てディスクアダプタDKA1またはDKA2と通信可能 である。以上、チャネルD11、D21の接続形態を例 ックループであり、ディスクドライブDK1、DK2、
- に説明したが、チャネルD12、D13、D14、D2 2、D23、D24でも同様である。本実施の形態のデ は、ディスクグループR1、R2、R3、R4を、それ のディスクドライブでRAIDグループを構成している が、RAIDグループを構成するドライブ数を4個に暇 ぞれRAIDグループとする。本実施の形態では、4個 **イスク装置においてRAIDシステムを構築する場合** るものではない。

クアレイDA1にアクセスできる。逆に、スイッチSW [0034] ディスクアレイDA1内の全ディスクドラ イブは、ディスクアダプタDKA1およびDKA2のど ちらからでもアクセス可能である。本実施の形態のディ イッチSW1が故障した場合でも、ディスクアダプタD KA1はチャネルD1bとスイッチSW2艇由でディス 1、SW2故障時の迂回経路として使用する。例えばス スク装置は、チャネルD1b、D2bをスイッチSW 20

ルアダプタCHA1と、キャッシュメモリCM1と、デ

ヲDKC2は、チャネルアダプタCHA2と、キャッシ

ィスクアダプタDKA1からなる。 ディスクコントロー

ネルD1aで接続し、ディスクアダプタDKA2とスイ

る。ディスクアダプタDKA1とスイッチSW1をチャ ュメモリCM2と、ディスクアダプタDKA2からな

ネルD 2 b とスイッチ S W 1 経由でディスクアレイDA 2が故障した場合は、ディスクアダプタDKA2はチャ | にアクセスできるので、信頼性が高いディスク装置が [0035] 図12に本発明の、第4の実施の形態であ るディスク装置の構成を示す。本実施の形態のディスク イッチSW1、SW2間を接破するチャネルD3a、D ディスクアダプタDKA1, DKA2と, スイッチSW アレイDA1からなる。スイッチSW1とSW2は、第 装置は、第3の実施の形態のディスク装置に対して、ス 1、SW2と、ディスクアレイDA1との間のデータ転 送方式は、ファイバチャネルを使用している。本実施の 1の実施の形態と同様に速度変換と多貫化、逆多重化を **行うスイッチとして機能する。ディスクコントローラD** 1スクコントローラDKC2は、チャネルアダプタCH 1. DKC2Ł, スイッチSW1, SW2Ł, ディスク モリCM1と、ディスクアダプタDKA1からなる。デ KC1は、チャネルアダプタCHA1と、キャッシュメ A2と、キャッシュメモリCM2と、ディスクアダプタ DKA2からなる。ディスクアダプタDKA1とスイッ チSW1をチャネルD1aで接続し、ディスクアタブタ 3 bを散けた点に特徴が有る。本実施の形態において、 形態のディスク装置は、ディスクコントローラDKC DKA2とスイッチSW2をチャネルD2aで接続し、

ディスクアダプタDKA1とスイッチSW2をチャネル SW1をチャネルD2bで接続する。さらに、スイッチ るディスクアレイと、D14とD24に接続した4個の D 1 b で接続し、ディスクアダプタDKA 2 とスイッチ D22に接続した4個のディスクからなるディスクアレ D11, D12, D13, D14, D21, D22, D スクドライブを接続する。ディスクアレイDA1内の金 ディスクドライブは、ディスクアダプタDKA1および DKA2のどちらからでもアクセス可能である。本実筋 する場合は、ディスクグループR1、R2、R3、R4 [0036] ディスクアレイDA1を構成するディスク ドライブは、入出力ポートを2個有する。例えば、ディ スクドライプDK1, DK2, DK3, DK4は, チャ ネルD11およびD21の両チャネルと接続する。ディ スクアレイDA1は、チャネルD11とD21に接続し た4個のディスクからなるディスクアレイと、D12と イと、D13とD23に接続した4個のディスクからな ディスクからなるディスクアレイ、からなる。 チャネル 23. D24は、図11に示すようにFC-ALでディ の形態のディスク装置においてRAIDシステムを構築 を、それぞれRAIDグループとする。本実施の形態で は、4個のディスクドライブでRAIDグループを構成 しているが、RAIDグループを構成するドライブ数を SW1とSW2をチャネルD3a、D3bで接続する。

イスクアダプタDKA1は、チャネルD1aとSW1を 介してディスクアレイDA1にアクセスする経路 (艇路 1) と、チャネルロ1 bとスイッチ SW2 とチャネルロ 3 a とスイッチ S W 1 を介してディスクアレイDA 1 に アクセスする経路(経路2)を有する。同僚に、ディス クアダプタDKA2は、チャネルD2aとSW2を介し てディスクアレイDA1にアクセスする経路(低路3) **常時(スイッチ故障無しの場合)について説明する。** イスクアレイDA1のアクセス経路について、先ず、 9

S

将開2003-3030

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

- セスする経路 (経路4)を有する。一方、スイッチ故障 と、チャネルD2bとスイッチSW1とチャネルD3b とスイッチSW2を介してディスクアレイDA1にアク 時は、チャネルD1b、D2bを迂回経路として使用す る。例えばスイッチSW1が枚降した場合でも、ディス クアダプタDKA1はチャネルD16とスイッチSW2 スイッチSW2が故障した場合は、ディスクアダプタD KA2はチャネルD2bとスイッチSW1뚎由でディス 紙由でディスクアレイDA1にアクセスできる。逆に、 クアレイDA1にアクセスできる。
- [0038]次に、本実施の形態におけるディスクアダ ブターディスクアレイ間のスルーブットについて説明す b、D3a、D3b上のデータ伝送遊取をチャネル当り 2Gbpsとし、チャネルD11, D12, D13, D 14, D21, D22, D23, D24上のデータ伝送 き、スイッチSW1とディスクアレイDA1回の総スル る。例として、チャネルDla、Dlb、D2a、D2 ープットは4Gbpsである。ディスクアダプタDKA 速度をチャネル当り1Gbpsであるとする。このと
 - クセスすることにより、総スループットは4Gbpsと と、ディスクアレイDA1個のスループットが共に4G 1とスイッチSW1間は、上紀経路1および経路2でア bpsであるので、ディスクアダプタDKA1とディス なる。スイッチSW1のディスクアダプタDKA1伽 クアレイDA1間のスループットは4Gbpsとなる。
 - 同様に、スイッチSW2とディスクアレイDA1間の総 スループットは4Gbpsである。ディスクアダプタロ でアクセスすることにより、総スループットは4Gbp KA2とスイッチSW2間は、上記種路3および種路4 sとなる。スイッチSW2のディスクアダプタDKA2 スクアレイDA2回のスルーブットは4Gbpsとなる 第3の実施の形態 (図8) において、上記のチャネル当 りスループット値を適用すると、チャネルD1b、D2 ディスクアダプタDKA1とディスクアレイDA1間の **密と、ディスクアレイDA1億のスループットが共に4** Gbpsであるので、ディスクアダプタDKA2とディ スループットは、チャネルD18上のスループットに飼 限され、2Gbpsとなる。同様に、ディスクアダプタ **しをスイッチ故障時の迂回経路としか使用しないので、**
 - チャネルD2a上のスループットに怠慢され、2Gbp DKA2とディスクアレイDA1間のスループットは、 20
 - [0037] ディスクアダプタDKA1, DKA2とデ

ら、本史筋の形態によれば、ディスクアダプタースイッ ターディスクアレイ間スループットを4Gbpsにする ためには、チャネルD1点およびD28のデータ伝送渡 **度を、それぞれ4Gbpgに高める必要がある。以上か チ間のチャネル当りデータ伝送速度が低くても、ディス** クアダプターディスクアレイ間の総スループットが高い 8 となる。第3の実施の形態において、ディスクアダプ ディスク鞍団が奥現できる。

Fの効果がある。ディスクアダプタとディスクアレイ関 ク数置を提供できる。また、信頼性の高いディスクアレ (発明の効果) 以上税明したように、本発明によれば以 た、ディスクアダプタとディスクアレイ間のスルーブッ レイ間のスループットが高く低コストなディスク装置を イを有するディスク装置を提供できる。また、信頼性が 商いディスクアダプタとディスクアレイ闘ネットワーク を存するディスク藝囮を提供できる。また、俗類性およ びスルーブットが高いディスクアダプタとディスクアレ できる。また、高スルーブットを維持できるディスク藝 トが高く、且つディスクドライブ接続台数が多いディス 5. また、ディスクからの脱み出しおよびディスクへの **掛きこみを高スループット化できるディスク装座を提供** 回を抱供できる。また、ディスクアダプタとディスクア のスループットが高いディスク数個を提供できる。虫 イ間ネットワーク、を有するディスク毎囮を提供でき [0039] 指供できる。

【図問の簡単な説明】

【図1】本発明の第1の実施の形態のディスク装置を示 す図である。

[図2] 本発明に用いるスイッチの構成を示す図であ 【図3】 本発明に用いるスイッチの構成を示す図であ

【図4】本発明に用いるスイッチの動作を示す図であ

【図5】本兜明に用いるスイッチの動作を示す図であ

[図6] 本発明期1の実施の形態に対して、ディスクド

【図7】本発明の第2の実施の形態のディスク装置を示 打図である。 [図8] 本発明の第3の実施の形態のディスク装置を示 一図である。

【図10】FC-ALによる接続形態を説明する図であ [図9] 従来のディスク装置を示す図である。

【図11】FC-ALによる接続形態を説明する図であ

【図12】本発明の第4の実施の形態のディスク装置を

示す図である。 [符号の説明] DKC, DKC1, DKC2 ディスクコントローラ CHA、CHA1、CHA2 チャネルアダプタ DKA, DKA1, DKA2 74307979 CM, CM1, CM2 キャッシュメモリ DA. DA1~DA4 F1X97V1

DK1~DK4 F1XDF517 R1~R6 ディスクグループ

(図3)

C1~C4, D1~D4, D01~D05, D11~D 14, D21~D24, D1a, Dib, D2a, D2 b. D3a, D3b チャネル

P1~P5 スイッチボート SW1~SW4 A1"F 25 XSW クロスパスイッチ

out1~out5 クロスパスイッチ出力ボー in1~in5 クロスパスイッチ入力ポート **CTL スイッチコントローラ**

ENC1、ENC2 8B10B変換エンコーダ DEC1, DEC2 8B10B変換デコーダ パラレルシリアル変換装置 SP1、SP2 シリアルパラレル変換装置 BM1. BM2 パッファメモリ T1, T2, T3 PS1, PS2

T4、T5、T6 フレームの時間 T× 泌癌癌

(図()

[885]

Rx 受信機

NL NL 法ート ライブを増設する方法を示す図である。

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

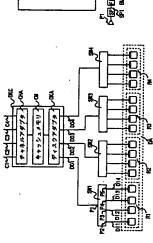
S

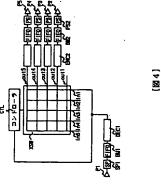
特開2003-30305

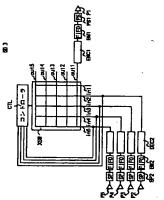
(図1)

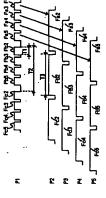
[図2]

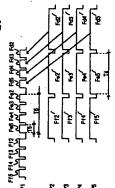
2

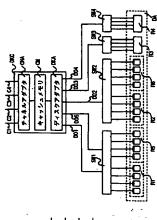












- 11 -

特開2003-303055

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

特開2003-30305

